On the Data Processing Theorem in the Semi–Deterministic Setting

Neri Merhav

Department of Electrical Engineering

Technion—Israel Institute of Technology

Haifa 32000, Israel

ITA 2014, San Diego, CA, February 2014.

Background

During the 70s and 80s, Ziv (partly, with Lempel) has developed the individual–sequence approach:

- 1978: Fixed–rate, almost–lossless compression using FSM's.
- 1978: The LZ algorithm.
- 1980: Lossy compression for noiseless/noisy transmission.
- 1984: Almost lossless comp. with side info @ decoder (Slepian–Wolf).
- 1986: Lossless compression of 2D data using FSM's.

Background

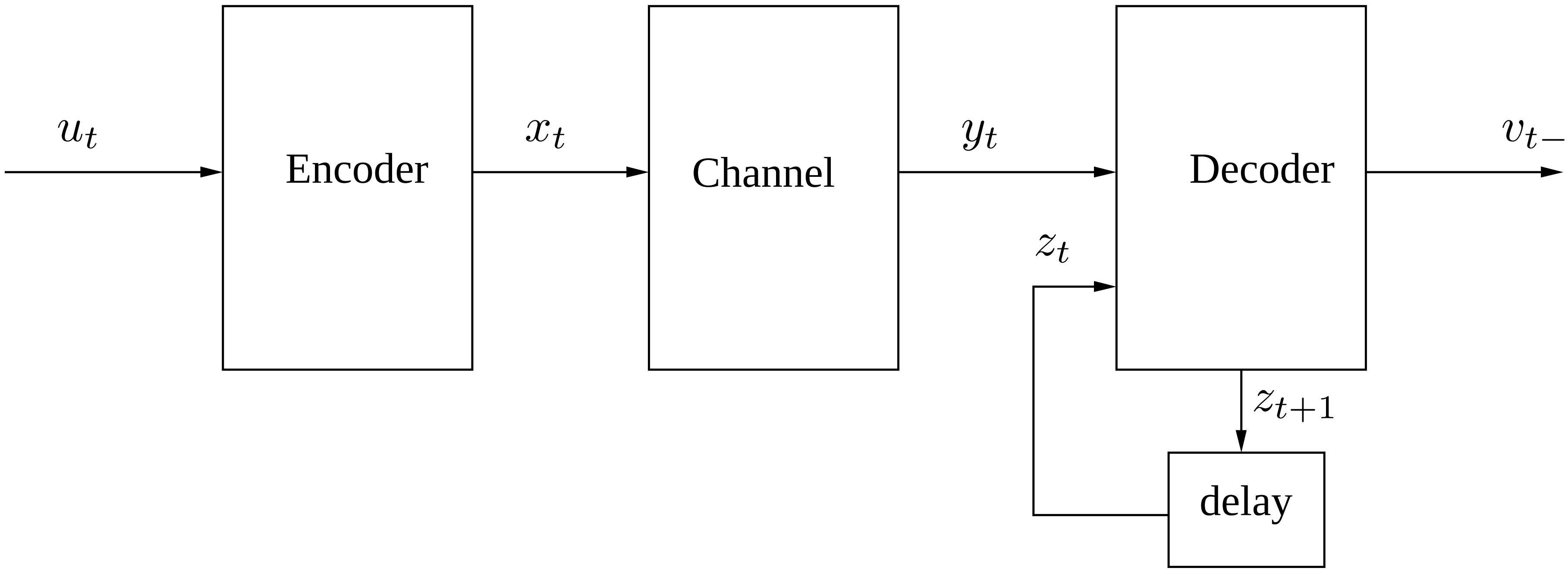
During the 70s and 80s, Ziv (partly, with Lempel) has developed the individual–sequence approach:

- 1978: Fixed–rate, almost–lossless compression using FSM's.
- 1978: The LZ algorithm.
- 1980: Lossy compression for noiseless/noisy transmission.
- 1984: Almost lossless comp. with side info @ decoder (Slepian–Wolf).
- 1986: Lossless compression of 2D data using FSM's.

Objectives

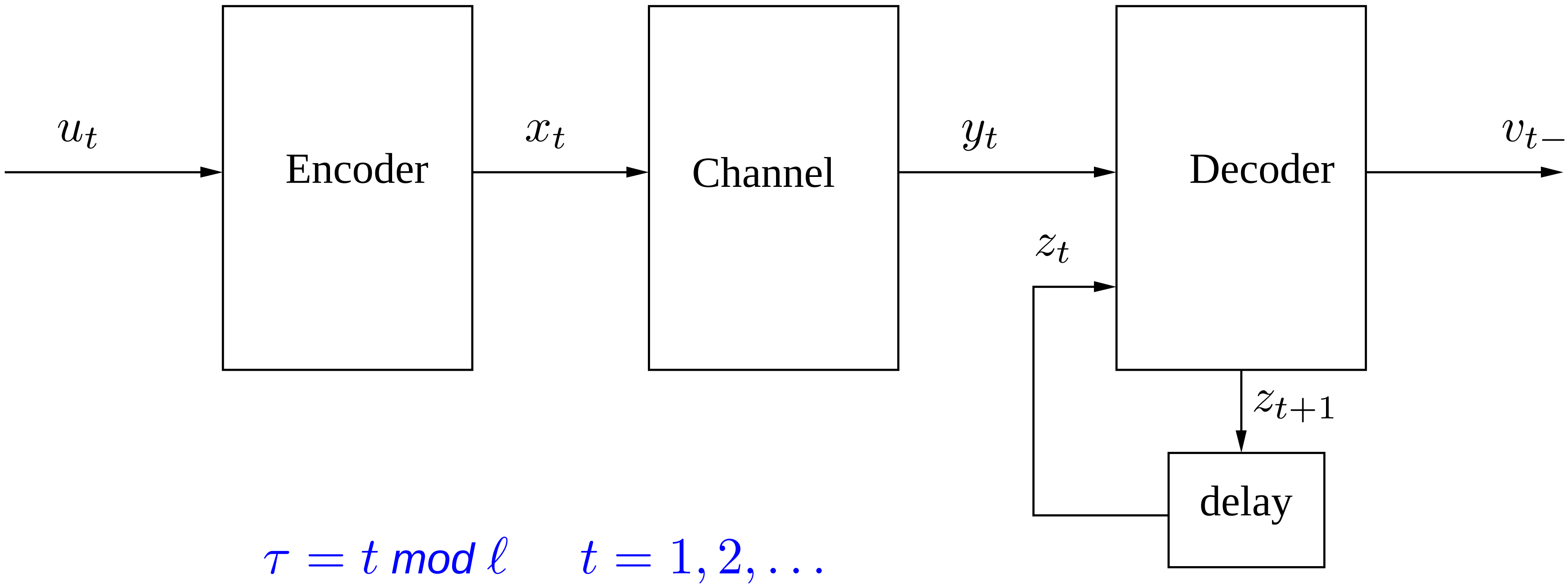
- Data processing thm of [Ziv80] revisited – the semi–deterministic setting:
  1. Correcting a few imprecise points.
  2. Strengthening the lower bound:
     - No limitations on the encoder (not necessarily an FSM).
     - Allowing a modulo–ℓ counter (periodically time–varying decoder).
- Tighter bound under a common reconstruction constraint.
- Bounds that depend on the LZ complexity of the input/output sequence.
- Analogous results for linear encoders/decoders in the continuous case.

The Communication System



$$v_{t-d} = f(z_t, y_t), \qquad t = d+1, d+2, \ldots$$
$$z_{t+1} = g(z_t, y_t), \qquad t = 1, 2, \ldots$$

The Communication System



$\tau = t \text{ mod } \ell \qquad t = 1, 2, \ldots$

$$v_{t-d} = f_\tau(z_t, y_t), \qquad t = d+1, d+2, \ldots$$
$$z_{t+1} = g_\tau(z_t, y_t), \qquad t = 1, 2, \ldots$$

Empirical Statistics Defined

Let ℓ divide n and consider the partition of $\boldsymbol{u} = (u_1, \ldots, u_n)$ into n/ℓ ℓ–blocks

$$\boldsymbol{u}_i = (u_{i\ell+1}, \ldots, u_{i\ell+\ell}), \qquad i = 0, 1, \ldots, n/\ell - 1$$

and similarly for other vectors (with $v_{n-d+1}, \ldots, v_n$ taking arbitrary values in $\mathcal{V}$).

We define $\hat{P}_{U^\ell X^\ell Y^\ell V^\ell Z}$ as the empirical joint distribution generated by counting the relative frequency of

$$\{\boldsymbol{u}_i = u^\ell,\ \boldsymbol{x}_i = x^\ell,\ \boldsymbol{y}_i = y^\ell,\ \boldsymbol{v}_i = v^\ell,\ z_{i\ell+1} = z\}, \qquad i = 0, 1, \ldots, n/\ell - 1$$

for all $u^\ell \in \mathcal{U}^\ell$, $x^\ell \in \mathcal{X}^\ell$, $y^\ell \in \mathcal{Y}^\ell$, $v^\ell \in \mathcal{V}^\ell$, and $z \in \mathcal{Z}$.

Some Concerns About the DPT in [Ziv80]

- Joint distribution defined s.t.
  - $U^\ell \perp (Z, Z')$ ($Z' =$ encoder state), but Z' responds to U^ℓ.
  - Y^ℓ is independent of Z' given X^ℓ, but Z responds to Y^ℓ.
- DPT: $U^\ell \perp V^\ell$ given (Z, X^ℓ), but $\exists$ empirical dependencies.
- v_{t-d} is renamed as v_t, but:
  - DPT applies to $\{u_t, v_{t-d}\}$
  - distortion is defined between $\{u_t, v_t\}$.
  - Consequence: Lower bound is independent of d.

Distortion Bound Without Common Reconstruction

Since $V^{\ell-d}$ is a deterministic function of Y^ℓ and Z,

$$\hat{I}(U^\ell; V^{\ell-d}|Z) \leq \hat{I}(U^\ell, X^\ell; Y^\ell|Z) \leq \hat{H}(Y^\ell) - \hat{H}(Y^\ell|U^\ell, X^\ell) + \log s$$

but on the other hand,

$$\hat{I}(U^\ell; V^{\ell-d}|Z) \geq \hat{I}(U^\ell; V^\ell) - \log s - d \log M.$$

Taking expectations and using $\mathbf{E}\hat{H}(Y^\ell|U^\ell, X^\ell) \sim H(Y^\ell|X^\ell)$,

$$\text{upper bound} \leq \ell C + \log s,$$

whereas

$$\text{lower bound} \geq \ell \cdot \hat{R}_{U^\ell}(\mathbf{E}\rho(U^\ell, V^\ell)) - \log s - d \log M.$$

Bound Without Common Reconstruction (Cont'd)

Theorem 1:

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{E}\{\rho(u_t,V_t)\} \geq \hat{D}_{U^\ell}\left(C + \frac{2\log s + d\log M}{\ell} + \delta_1(\ell,n)\right)$$

where

$$\delta_1(\ell,n) = \frac{(JK)^\ell \log L}{\sqrt{n}} + \frac{(JKL)^\ell \log e}{2n} + o\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\hat{D}_{U^\ell}(R) = \min\left\{\frac{\mathbf{E}\rho(U^\ell,\tilde{V}^\ell)}{\ell} \,:\, \hat{I}(U^\ell;\tilde{V}^\ell) \leq \ell R\right\}.$$

- Blue term $=$ "effective extra capacity" due to memory and allowed delay.
- In the absence of a mod–ℓ counter, maximize the r.h.s. over ℓ.
- Dependence on ℓ – complicated.

Alternative Bound for Difference Distortion Measures

Assume $\rho(u, v) = \varrho(v - u)$, and define

$$\Phi(D) = \max\{H(W) : \mathbf{E}\rho(W) \leq D\}$$

$$\Psi(t) = \begin{cases} \Phi^{-1}(t) & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Theorem 2:

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{E}\{\varrho(V_t - u_t)\} \geq \Psi\left(\frac{c(u^n)\log c(u^n)}{n} - C - \frac{2\log s + d\log M}{\ell} - \delta_2(\ell, n)\right),$$

where $c(u^n)$ is the number of LZ phrases of u^n and

$$\delta_2(\ell, n) = \delta_1(\ell, n) + \frac{2\ell(1 + \log J)^2}{(1 - \epsilon_n)\log n} + \frac{2\ell J^{2\ell}\log J}{n} + \frac{1}{\ell},$$

Common Reconstruction

For a given $\epsilon_n \to 0$, suppose that $\exists\ q$ s.t.

$$\mathbf{E}\hat{\text{Pr}}\{V^\ell \neq \hat{V}^\ell\} \equiv \frac{\ell}{n} \sum_{i=0}^{n/\ell-1} \text{Pr}\{V_{i\ell+1}^{i\ell+\ell} \neq \hat{v}_{i\ell+1}^{i\ell+\ell}\} \leq \epsilon_n,$$

where $\hat{V}^\ell = q(U^\ell)$.

Theorem 3:

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{E}\{\rho(u_t, V_t)\} \geq \tilde{D}_{U^\ell}\left(C + \frac{2\log s + d\log M}{\ell} + \delta_2(\ell, n) + 2\Delta(\epsilon_n)\right) - \rho_{\max}\epsilon_n,$$

where $\Delta(\epsilon_n) = h_2(\epsilon_n) + \epsilon_n \ell \log J$, $h_2(\cdot)$ being the binary entropy function, and

$$\tilde{D}_{U^\ell}(R) = \min_q \left\{\frac{1}{\ell}\hat{\mathbf{E}}\rho(U^\ell, q(U^\ell)) :\ \hat{H}(q(U^\ell)) \leq \ell R\right\}.$$

$\hat{H}(q(U^\ell))$ can be further lower bounded in terms of $c(\hat{v}^n)\log c(\hat{v}^n)$.

Linear Encoders/Decoders

Analogous results can be obtained in the linear/Gaussian/quadratic case.

Let $\rho(u, v) = (u - v)^2$ and consider a linear encoder

$$x_t = \sum_{i=1}^{\infty} a_i x_{t-i} + \sum_{i=0}^{\infty} b_i u_{t-i},$$

a Gaussian memoryless channel

$$Y_t = x_t + N_t, \qquad N_t \sim \mathcal{N}(0, \sigma^2),$$

and a linear decoder defined by

$$\begin{aligned} v_{t-d} &= \alpha z_t + \beta y_t \\ z_{t+1} &= \gamma z_t + \delta y_t. \end{aligned}$$

Linear Encoders/Decoders (Cont'd)

Main idea: Assume (for a moment) a Gaussian input process:

- All processes in the system are jointly Gaussian.
- All information measures depend solely on auto/cross–correlations.
- DPT gives inequality relations between these auto/cross–correlations.
- Relations between auto/cross–correlations do not rely on Gaussianity.
- The input does not have to be Gaussian.
- Use the empirical cov. matrix of U^ℓ to derive semi–deterministic bounds.
- Due to the linear–Gaussian structure, the test channel is $V^\ell = GU^\ell + W^\ell$.

Linear/Gaussian Lower Bound

Theorem 4:

$$\frac{1}{n}\sum_{t=1}^{n}\mathbf{E}(V_t - u_t)^2 \geq \hat{D}\left(C + \frac{1}{\ell}\log\frac{\sigma_Z^2}{\epsilon_Z^2} + \frac{d}{2\ell}\log\frac{\sigma_V^2}{\epsilon_V^2} + \epsilon_n\right)$$

where

$$\hat{D}(R) = \min_{G,\Sigma_W}\Big\{\mathrm{tr}\{(G-I)\hat{\Sigma}_U(G^T - I) + \frac{1}{\ell}\Sigma_W\} : \frac{1}{2\ell}\log|I + \Sigma_W^{-1}G\hat{\Sigma}_U G^T| \leq R\Big\}$$

and

$$\hat{\Sigma}_U = \frac{\ell}{n}\sum_{i=0}^{n/\ell-1}\boldsymbol{u}_i\boldsymbol{u}_i^T.$$